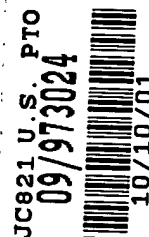


IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of :
Genichiro SUEKI et al. :
Serial No. NEW : Attn: APPLICATION BRANCH
Filed October 10, 2001 : Attorney Docket No. 2001_1332A



INTERESTED ARTICLE SERVING
SYSTEM AND INTERESTED ARTICLE
SERVING METHOD

CLAIM OF PRIORITY UNDER 35 USC 119

Assistant Commissioner for Patents,
Washington, DC 20231

Sir:

Applicants in the above-entitled application hereby claim the date of priority under the International Convention of Japanese Patent Application No. 2000-351144, filed November 17, 2000, as acknowledged in the Declaration of this application.

A certified copy of said Japanese Patent Application is submitted herewith.

Respectfully submitted,

Genichiro SUEKI et al.

By Charles R. Watts
Charles R. Watts
Registration No. 33,142
Attorney for Applicants

CRW/asd
Washington, D.C. 20006-1021
Telephone (202) 721-8200
Facsimile (202) 721-8250
October 10, 2001

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

JC821 U.S. PTO
09/973024
10/10/01

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年11月17日

出 願 番 号

Application Number:

特願2000-351144

出 願 人

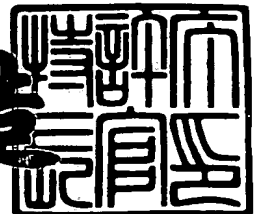
Applicant (s):

三菱スペース・ソフトウェア株式会社

2000年12月15日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造



出証番号 出証特2000-3105057

【書類名】 特許願
【整理番号】 00-0129
【あて先】 特許庁長官殿
【国際特許分類】 G06F 17/30
【発明者】

【住所又は居所】 神奈川県鎌倉市上町屋 5 2 4 番地 三菱スペース・ソフトウエア株式会社鎌倉事業部内

【氏名】 末木 源一郎

【発明者】

【住所又は居所】 神奈川県鎌倉市上町屋 5 2 4 番地 三菱スペース・ソフトウエア株式会社鎌倉事業部内

【氏名】 藤木 宏明

【発明者】

【住所又は居所】 神奈川県鎌倉市上町屋 5 2 4 番地 三菱スペース・ソフトウエア株式会社鎌倉事業部内

【氏名】 吉野 直子

【特許出願人】

【識別番号】 591102095

【氏名又は名称】 三菱スペース・ソフトウェア株式会社

【代理人】

【識別番号】 100104776

【弁理士】

【氏名又は名称】 佐野 弘

【手数料の表示】

【予納台帳番号】 053246

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0016932

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 興味記事配信システム及び興味記事配信方法

【特許請求の範囲】

【請求項 1】 クライアント装置と該クライアント装置からの要求に応じて情報を提供できるサーバ装置とがネットワーク接続されている記事配信システムにおいて、

前記サーバ装置は、前記クライアント装置からの要求を受けこれに応答するWWWサーバ部と、前記クライアント装置を使用する顧客の個人情報及びアクセス履歴を保存するデータベースサーバ部と、該データベースサーバ部に保存された前記顧客のアクセス履歴を分析して検索条件式を生成する興味記事抽出サーバ部と、該興味記事抽出サーバ部で生成された検索条件式に基づいて、外部の記事生成コンピュータから逐次送出される記事データを検索する超並列計算機とからなり、

前記超並列計算機の複数の異なるプロセッサ上に複数の異なる前記検索条件式を別個に設定し、前記記事データを全文検索して前記検索条件式に合致した結果を前記クライアント装置に送信するようにしたことを特徴とする興味記事配信システム。

【請求項 2】 前記記事データは、前記サーバ装置とは別個の外部の記事生成コンピュータから受信したものであることを特徴とする請求項 1 に記載の興味記事配信システム。

【請求項 3】 前記顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として内容が類似する複数の記事を複数回に渡って前記顧客がアクセスした場合に、前記内容が類似する複数の記事に含まれる自然語を前記検索条件式の検索キーワードとして使用するようにしたことを特徴とする請求項 1 又は 2 に記載の興味記事配信システム。

【請求項 4】 前記顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として内容の異なる複数の記事に前記顧客がアクセスした場合に、一定記事数又は一定期間にアクセスした記事数を基準記事数と定め、該基準記事数における前記内容の異なる複数の記事に現れる同一の自然語を前記検索条件式の検索キーワー

ドとして使用することを特徴とする請求項 1 又 2 に記載の興味記事配信システム

【請求項 5】 前記顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として前記顧客が特定の記事データを指定した場合に、該特定の記事データに含まれる自然語を前記検索条件式の検索キーワードとして使用するようにしたことを特徴とする請求項 1 又は 2 に記載の興味記事配信システム。

【請求項 6】 前記検索条件式は、前記顧客が指定した特定の記事データにおける自然語の出現回数と前記顧客のアクセス履歴中に含まれる複数の記事中に前記自然語が出現する割合とによる重み付けをした特徴的な自然語を前記検索キーワードに採用したことを特徴とする請求項 4 に記載の興味記事配信システム。

【請求項 7】 前記クライアント装置が携帯端末装置であることを特徴とする請求項 1 乃至 6 のいずれか 1 つに記載の興味記事配信システム。

【請求項 8】 クライアント装置と該クライアント装置からの要求に応じて情報を提供できるサーバ装置とがネットワーク接続されている場合の記事配信方法において、前記クライアント装置を使用する顧客のアクセス履歴を分析して複数の異なる検索条件式を生成し、該複数の異なる検索条件式を複数の異なるプロセッサを有する超並列計算機の前記複数の異なるプロセッサ上に別個に設定し、外部の記事生成コンピュータから逐次送出される記事データを前記複数の異なる検索条件式で同時並行的に全文検索し、前記検索条件式に合致した結果を前記クライアント装置に提供するようにしたことを特徴とする興味記事配信方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、ネットワーク上の記事配信サービスにおいて、顧客が記事の配信を受ける際、顧客の興味の対象となる記事を自動的に選定し、記事の入手してから即時に、顧客の興味ある記事のみを配信することができるようにした興味記事配信システムと興味記事配信方法に関するものである。

【0002】

【従来の技術】

従来、インターネットを通じて、多種多様な記事配信サービスが運営されているが、配信元による編集若しくは分類のための処理等を経て配信されるのが一般的である。

【0003】

特に、一般的なデータベースシステムを有する計算機では、検索対象とする記事をデータベースに投入する際、記事検索時のキーワードとなり得る特徴のある語をインデックスとして付与する処理（インデックス付け）を事前準備として行う必要があるため、新しい記事が発生してから、即時に記事を配信することは困難であった。

【0004】

顧客が興味ある記事を得るための一般的なサービスは、顧客が指定したカテゴリ、及びキーワードにしたがって、そのカテゴリに含まれる記事又はそのキーワードを含む記事を検索し、その検索条件に合致した記事を電子メール等の配信手段を用いて、配信するという形態がとられている。しかし、カテゴリ単位での指定では、カテゴリ内のすべての記事が配信されてしまうため、顧客の希望しない不必要な記事を含むことが多くなる、或いは、顧客の興味が複数に渡る場合、非常に多くの記事が配信されてしまう。

【0005】

また、顧客がキーワードを指定する場合は、そのキーワードを含む記事がすべて配信され、顧客の興味と関係のない記事が配信されてしまう。或いは、キーワードの設定に不慣れな顧客は、所望の記事を取得することが非常に困難であった。

【0006】

さらに、顧客の興味をアクセス履歴等に基づき、様々な手法により分析する技術が考案されつつあるが、これまでの手法は顧客の興味が定性的なものという前提に基づくものであり、時間経過による興味の移り変わりに対応するものではなかった。

【0007】

リアルタイムに知りたい興味のほとんどは、現在世間を賑わしている事件の経

過や期間限定のイベント（例えば、オリンピック等）の場合が多く、これらは時間の経過と共に興味も薄れてくるのが一般的である。

【0008】

現に、記事の参照時間、クリック回数をパラメータとして基本的にこれらの回数が多いものが、最も興味があるという仕組みが考案されているが、アクセス履歴が長期間に渡っても、最近興味ある記事の参照時間やクリック回数が過去に興味があったものを上回る回数になるとは限らない。

【0009】

また、アクセス履歴に基づく顧客の興味を表すキーワードの自動抽出手法として、当該キーワードの記事ごとの出現頻度とすべての記事中に当該キーワードが出現する割合（重要度の大きい）を算出し、これを掛け合わせたものを重みとする手法もあるが、これでは記事によって偏った傾向が出力される可能性がある。例えば、たまたま1回アクセスした記事に重要度の大きいキーワードが複数回出現していた場合は、興味がない記事であっても興味があったものとみなされてしまう。

【0010】

【発明が解決しようとする課題】

そこで、この発明では、顧客によるカテゴリやキーワードの設定が不要で、顧客が意識することなく、顧客の興味を自動的に抽出し、即時に顧客の興味のある記事を配信するシステムとその方法を提供することを課題とする。

【0011】

また、顧客が、自分の興味を指定する方法においても、顧客は、特に興味のある情報について、キーワードで指定するのではなく、その情報が記述されている記事全体を指定することにより、顧客の興味に合った記事を的確にかつ即時に配信することが可能なシステムとその方法を提供することを課題とする。

【0012】

【課題を解決するための手段】

かかる課題を解決するために、請求項1に記載の発明は、クライアント装置と該クライアント装置からの要求に応じて情報を提供できるサーバ装置とがネット

ワーク接続されている記事配信システムにおいて、前記サーバ装置は、前記クライアント装置からの要求を受けこれに応答するWWWサーバ部と、前記クライアント装置を使用する顧客の個人情報及びアクセス履歴を保存するデータベースサーバ部と、該データベースサーバ部に保存された前記顧客のアクセス履歴を分析して検索条件式を生成する興味記事抽出サーバ部と、該興味記事抽出サーバ部で生成された検索条件式に基づいて、外部の記事生成コンピュータから逐次入力される記事データを検索する超並列計算機とからなり、前記超並列計算機の複数の異なるプロセッサ上に複数の異なる前記検索条件式を別個に設定し、前記記事データを全文検索して前記検索条件式に合致した結果を前記クライアント装置に送信するようにしたことを特徴としている。

【 0 0 1 3 】

請求項 2 に記載の発明は、請求項 1 の構成に加えて、前記記事データは、前記サーバ装置とは別個の外部の記事生成コンピュータから受信したものであることを特徴としている。

【 0 0 1 4 】

請求項 3 に記載の発明は、請求項 1 又は 2 の構成に加えて、前記顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として内容が類似する複数の記事を複数回に渡って前記顧客がアクセスした場合に、前記内容が類似する複数の記事に含まれる自然語を前記検索条件式の検索キーワードとして使用するようにしたことを特徴としている。

【 0 0 1 5 】

請求項 4 に記載の発明は、請求項 1 又は 2 の構成に加えて、前記顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として内容の異なる複数の記事にアクセスした場合に、一定記事数又は一定期間にアクセスした記事数を基準記事数と定め、該基準記事数における前記内容の異なる複数の記事に現れる同一の自然語を前記検索条件式の検索キーワードとして使用することを特徴としている。

【 0 0 1 6 】

請求項 5 に記載の発明は、請求項 1 又は 2 に構成に加えて、前記顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として前記顧客が特定の記事データを

指定した場合に、該特定の記事データに含まれる自然語を前記検索条件式の検索キーワードとして使用するようにしたことを特徴としている。

【 0 0 1 7 】

請求項 6 に記載の発明は、請求項 4 の構成に加えて、前記検索条件式は、前記顧客が指定した特定の記事データにおける自然語の出現回数と前記顧客のアクセス履歴中に含まれる複数の記事中に前記自然語が出現する割合とによる重み付けをした特徴的な自然語を前記検索キーワードに含めたことを特徴としている。

【 0 0 1 8 】

請求項 7 に記載の発明は、請求項 1 乃至 6 のいずれか 1 つの構成に加えて、前記クライアント装置が携帯端末装置であることを特徴としている。

【 0 0 1 9 】

請求項 8 に記載の発明は、クライアント装置と該クライアント装置からの要求に応じて情報を提供できるサーバ装置とがネットワーク接続されている場合の記事配信方法において、前記クライアント装置を使用する顧客のアクセス履歴を分析して複数の異なる検索条件式を生成し、該複数の異なる検索条件式を複数の異なるプロセッサを有する超並列計算機の前記複数の異なるプロセッサ上に別個に設定し、外部の記事生成コンピュータから逐次送出される記事データを前記複数の異なる検索条件式で同時並行的に全文検索し、前記検索条件式に合致した結果を前記クライアント装置に提供するようにしたことを特徴としている。

【 0 0 2 0 】

【発明の実施の形態】

この発明の実施の形態に係る興味記事配信システムの構成について、図 1 及び図 2 に従って説明する。

【 0 0 2 1 】

図 1 は、この発明の実施の形態に係る興味記事配信システムの構成を示すブロック図である。図 2 は、この発明の実施の形態に係る興味記事配信システムのサーバ装置に含まれる超並列計算機の内部処理の仕組みを示した模式図である。

【 0 0 2 2 】

図 1 において、サーバ装置 1 は、興味記事配信システムの統括管理を行うコン

ピュータであり、WWWサーバ部 2、電子メールサーバ部 3、データベースサーバ部 4、興味記事抽出サーバ部 5 及び超並列計算機 6 から構成されている。

【 0 0 2 3 】

クライアント装置 7 は、顧客によって使用されるインターネット 8 に接続されるコンピュータであり、顧客はホームページの閲覧、興味記事の指定、電子メールの閲覧をこのクライアント装置 7 により行う。クライアント装置 7 は、インターネット 8 に接続される携帯電話等の携帯端末装置又は電子メールサービスが受けられる携帯電話等の携帯端末装置であってもよい。

【 0 0 2 4 】

記事生成コンピュータ 9 は、新規に発生した記事を外部の記事配信機関がサーバ装置 1 の WWWサーバ部 2 に登録するためのコンピュータであり、新規に発生した記事は同時にサーバ装置 1 の超並列計算機 6 及び／又はデータベースサーバ部 4 へも転送される。

【 0 0 2 5 】

興味記事抽出サーバ部 5 は、データベースサーバ部 4 とこのデータベースサーバ部 4 に接続されている超並列計算機 6 とに接続され、顧客のアクセス履歴若しくは顧客が指定した興味記事から検索条件式を自動的に生成する機能を有している。

【 0 0 2 6 】

超並列計算機 6 は、図 2 に示したように、数千～数万のプロセッサ 1 0（以下、これらをまとめてパイプラインという。）を内蔵することにより、このパイプライン 1 1 に複数の異なった検索条件式 1 2 を同時に設定可能としている。そして、これら大量のプロセッサ 1 0 を同時に動作させることによって、記事生成コンピュータ 9 から新たな記事 1 3 が送信されてくると、パイプライン 1 1 に記事データ 1 4 を送出し複数の異なった検索条件式 1 2 と記事データ 1 4 のマッチングを行う全文検索を実行する。マッチングの結果、検索条件式 1 2 に合致する記事データ 1 4 h が見つかったら、その記事データ 1 4 h がヒットしたとみなす機能を有する。

【 0 0 2 7 】

超並列計算機 6 は、全文検索エンジン（例えば、P a r a c e l 社製、F D F（登録商標） 4 T T e x t F i n d e r）のような機器が望ましいが、これと同等の機能及び性能を有するワークステーション等の機器でもよい。

【 0 0 2 8 】

WWWサーバ部 2 は、超並列計算機 6 による全文検索の結果得られた記事データ 1 4 h を顧客が指定する配信先アドレス若しくは W e b ページに転送する機能を有する。

【 0 0 2 9 】

データベースサーバ部 4 には、クライアント装置 7 を使用する顧客のアクセス履歴として、顧客ごとに直近の一定期間若しくは一定記事数において顧客がホームページを閲覧した履歴が保存されている。さらに、顧客ごとの興味度が高い自然語とその重要度を統計処理した結果等が併せて保存されてもよい。なお、アクセス履歴の領域節約のために、記事配信を要望する顧客と要望していない顧客について分割して定義し、要望していない顧客についてはホームページの閲覧履歴のみを保存してもよい。

【 0 0 3 0 】

記事データは、過去受信した記事を含む記事のタイトルと記事の本文であり、アクセス履歴の統計処理を行う際の自然語の選定に使用する。

【 0 0 3 1 】

さらに、データベースサーバ部 4 には、ユーザ情報としては、顧客の個人情報であるユーザー I D、ユーザ名、興味記事の配信先（U R L、メールアドレス等）が保存されている。

【 0 0 3 2 】

次に、この発明の実施の形態に係る興味記事配信システムの使用方法について、図 3 乃至図 5 に従って説明する。

【 0 0 3 3 】

図 3 は、この発明の実施の形態に係る興味記事配信システムを使用する場合の手順を示すフローチャートである。図 4 は、この発明の実施の形態に係る興味記事配信システムのクライアント装置の画面イメージ図である。

【0034】

まず、顧客は記事を閲覧する際、クライアント装置7をインターネット8に接続し、興味ある記事が掲載されているWebページ（ホームページ）にアクセスする（ステップS1）。顧客により記事が閲覧されると、サーバ装置1はWWWサーバ部2を介して顧客の個人情報とアクセス履歴を取得する。

【0035】

次に、取得した顧客の個人情報により、興味記事配信システムの利用を希望している顧客かどうかをユーザ情報との照会により調査する（ステップS2）。ユーザ情報には、項目に関する規定として、顧客を識別するための情報と、興味記事の配信先が必ず含まれ、それ以外に関しては、特に規定しない。

【0036】

調査の結果、興味記事配信システムの利用を希望する顧客であれば、アクセス履歴をデータベースサーバ部4へ登録する（ステップS3）。アクセス履歴には、顧客ごとに、アクセスした記事の記事全文、或いは、記事の所在を示す情報等、記事を特定できる情報、アクセスした日或いは日時、また関連記事指定がされているか否かの情報が含まれていればよい。

【0037】

記事の所在を、記事を特定する情報とする場合には、記事本文との対応情報が必要となる。アクセス履歴としては、記事本文に限られるものではなく、記事のタイトルや要約等でもよい。

【0038】

関連記事指定に関する情報については、アクセス履歴からの興味自動抽出機能のみで、関連記事指定機能を保有しない場合は、不要である。

【0039】

次に、顧客が閲覧した際に、アクセス履歴から、その記事が興味記事として指定されているかどうかを調査する（ステップS4）。関連記事指定機能を有しない場合は、閲覧内容の確認は不要である。

【0040】

閲覧内容の確認の結果、興味記事として指定されている場合、その記事内容が

ら、検索条件式12を自動生成する（ステップS6）。指定された記事内容から検索条件式12を自動生成する手順は、図5のフローチャートに示しているが、その詳細は後述する。

【0041】

閲覧内容の確認の結果、興味記事として指定されていない場合、アクセス履歴から統計処理を行ない、検索条件式12を自動生成（更新）する（ステップS7）。アクセス履歴の統計処理及びこれに基づく検索条件式12の自動生成（更新）の手順は、図6のフローチャートに示しているが、その詳細は後述する。

【0042】

検索条件式12の自動生成及びアクセス履歴の統計処理が済んだら、配信対象となる新規記事が、存在するかどうか調査する（ステップS8）。

【0043】

新規記事が存在する場合、顧客の検索条件式12と合致するかどうかを超並列計算機6により全文検索する（ステップS9）。新規記事が顧客の検索条件式12と合致しない場合には、次の新規記事が存在するかどうかの調査へ移る。

【0044】

新規記事が顧客の検索条件式12と合致した場合には、データベースサーバ部4のユーザ情報を参照し、顧客の指定する配信希望先がWebページ（ホームページ）か、電子メールかを調査する（ステップ10）。この時点では、顧客の使用するクライアント装置7には、図4（b）の下部に示したような配信希望先を指定する画面が表示されてもよい。

【0045】

調査の結果、配信希望先がWebページ（ホームページ）ならば、興味記事配信内容を反映したホームページを自動更新する（ステップS11）。配信希望先が電子メールならば、指定されているメールアドレスに興味記事配信内容を送信する（ステップS12）。ここで、興味記事を受信したクライアント装置7には、図4（c）に示したような興味記事表示画面が表れる。

【0046】

次に、顧客が興味記事を指定している場合の検索条件式12の自動生成の手順

について、図5に従って説明する。

【0047】

閲覧内容の確認の結果、閲覧した記事が興味記事として指定されている場合には、図3のフローチャートで示したとおり、その記事内容から、検索条件式12を自動生成することになる（ステップS6）。

【0048】

この場合、まず、顧客が指定した記事を形態素解析等により品詞単位に分割し、品詞情報を取得する（ステップ61）。次に、品詞単位に分割した自然語を、例えば、連続する名詞は結合させる等の複合語処理することにより、複合語を作成する（ステップ62）。複合語作成により、自然語の分割による表現の抽象化等を回避することができ、顧客の興味を的確に表す検索キーワードを選定することができる（ステップ63）。

【0049】

検索キーワードが選定されたらその検索キーワードをもとに、指定した記事内容に関連した文書を検索するための（検索キーワード、重み、スコアリング方式等からなる）検索条件式12を自動生成する（ステップS64）。

【0050】

スコアリング方式としては検索対象記事内に一度でも出現すればスコアを計上する方式や、検索対象記事内で、出現するたびにスコアを計上する方法等が挙げられる。

【0051】

以下に、前述の検索キーワードの選定に関する品詞情報、自然語が現れる文書数、自然語の一般性の選定条件について説明する。

【0052】

品詞情報から、複合語、名詞、未定義語等、特徴を表す品詞の自然語を抽出する。抽出対象とする品詞は、上記の複合語、名詞、未定義語に限られるものではなく、配信対象となる文書ごとに特徴を表すと予測される品詞を指定する。品詞単位に分割された後、場合によっては、必ず削除する自然語等をリストとして保有し、抽出対象品詞であっても、その自然語を削除する機能を追加してもよい。

【0053】

品詞情報により、絞り込まれた自然語に対し、自然語の出現回数や、自然語の一般性を表す重要度等からその自然語に対する重みを算出する。

【0054】

自然語の出現回数は、出現回数が多い自然語は、その記事の概念を表す自然語と考えられる。自然語の一般性を表す重要度は、例えば、様々な記事を格納したデータベースにおける、その自然語が出現する記事の割合を算出する。その割合が多ければ様々な記事に出現する自然語であるため、重要度が小さくなり、割合が少なければ、ある特定の記事にのみ現れる自然語として、重要度が大きくなる。出現頻度が多い自然語として、一般的な自然語も考えられるが、そのような自然語は、自然語の重要度が小さく、その自然語に付与される重みは、小さくなるため、検索結果に悪影響を与えることは少ない。

【0055】

次に、顧客が興味文書として指定していない場合の手順について、図6のフローチャートに従って説明する。

【0056】

閲覧内容の確認の結果、興味記事として指定していない場合には、図3のフローチャートで示したとおり、顧客のアクセス履歴から統計処理を行ない、検索条件式を自動生成することになる（ステップS7）。

【0057】

この場合、まず、顧客のアクセス履歴から、顧客のアクセス記事内容を抽出する（ステップS71）。アクセス履歴に記事本文が格納されている場合には、記事本文を抽出すればよい。また、アクセス履歴に格納されている情報が記事本文ではなく、記事の所在等、記事を特定するための情報である場合には、記事を特定するための情報と記事本文との対応情報から、記事内容を抽出する。

【0058】

次に、アクセス記事を形態素解析等により品詞単位に分割し、品詞情報を取得する（ステップS72）。そして、品詞単位に分割した自然語を、例えば、連続する名詞は結合させる等の複合語処理することにより、複合語を作成する（ステ

ップ S 7 3)。複合語の作成により、自然語の分割による表現の抽象化等を回避することができ、顧客の興味を適確に表す検索キーワードを選定する（ステップ S 7 4）。選定した検索キーワードをもとに、指定した記事内容に関連した記事を検索するための（検索キーワード、重み、スコアリング方式等からなる）検索条件式を自動生成（更新）する（ステップ S 7 5）。重みは、例えば、出現回数や自然語の重要性から算出する。スコアリング方式としては検索対象記事内に一度でも出現すればスコアを計上する方式や、検索対象記事内で、出現するたびにスコアを計上する方法等が挙げられる。

【 0 0 5 9 】

以下に、ステップ S 7 4 における検索キーワード選定に関する品詞情報、自然語が現れる文書数、自然語の一般性の選定条件について説明する。このうち、品詞情報の選定条件については、顧客が興味記事を指定している場合と同様であるためその説明を省略する。

【 0 0 6 0 】

品詞情報により、絞り込まれた自然語に対し、アクセス記事内において、複数の記事に現れる自然語は、顧客の興味を表す概念であるとして、所定の閾値を超える記事に現れた自然語のみ絞り込む。このとき、記事毎の自然語の出現頻度を考慮しないことが重要である。

【 0 0 6 1 】

次に、この絞り込まれた自然語を、さらに特徴的な自然語のみに絞り込むため、様々な記事で記述される自然語を除外する。そのために、例えば、様々な記事を格納したデータベースサーバ部 4 におけるその自然語が出現する記事の割合を算出し、ある所定の閾値を超える或いは下回る自然語については、一般的であると判断し、除外する。以上により絞り込まれた特徴的な自然語を顧客の興味を表す検索キーワードとすることで、よりノイズの少ない検索結果が得られる。

【 0 0 6 2 】

複数記事においては、記事ごとの表記ゆれ等が発生するため、上記の絞り込みにより、除外された自然語に関して、抽出した検索キーワードと部分一致する自然語に関しては、その自然語も検索キーワードとするという機能を追加してもよ

い。

【 0 0 6 3 】

上述した実施の形態では、インターネット 8 を接続手段としているが、WWWサーバ部 2 に接続可能な、特定回線等他のネットワークによる接続でもよい。

【 0 0 6 4 】

また、WWWサーバ部 2 は、顧客の情報閲覧及び閲覧の履歴が把握可能なシステム若しくは顧客からの興味記事の指定が可能なシステムにおいて顧客に提供する記事を一元的に管理する機能を有するものであればよく、必ずしもインターネットに常時接続されているものでなくともよい。

【 0 0 6 5 】

サーバ装置 1 の構成として説明したWWWサーバ部 2、電子メールサーバ部 3、興味記事抽出サーバ部 4 は、それぞれ独立した構成であってもよい。

【 0 0 6 6 】

また、サービスの形態によっては、WWWサーバ部 2、電子メールサーバ部 3 がいずれか若しくは両方が興味記事抽出サーバ部 4 と配置的に同じ場所に設置されていなくても、ネットワーク等で接続されている形態であればよい。

【 0 0 6 7 】

電子メールサーバ部 3 については、当該メール配信を行わない形態の場合は構成から除くこともできる。

【 0 0 6 8 】

アクセス履歴、記事内容、ユーザ情報は、それぞれ独立したデータベース管理システムなどの特定のソフトウェアにより管理されるデータベース若しくはデータベース管理システム等の特別なデータベース構造を持たないファイルによる構成であってもよい。

【 0 0 6 9 】

記事生成コンピュータ 9 は、同一ネットワーク上若しくは遠隔地からのファイル転送などにより、新規に発生した記事をサーバ装置 1 へ転送する機能を有するものであればよい。また、サーバ装置 1 と兼用してもよい。

【 0 0 7 0 】

【発明の効果】

以上説明したように、請求項 1 に記載の発明によれば、サーバ装置は、クライアント装置からの要求を受けこれに応答する WWW サーバ部と、クライアント装置を使用する顧客の個人情報及びアクセス履歴を保存するデータベースサーバ部と、該データベースサーバ部に保存された顧客のアクセス履歴を分析して検索条件式を生成する興味記事抽出サーバ部と、該興味記事抽出サーバ部で生成された検索条件式に基づいて、外部の記事生成コンピュータから逐次送出される記事データを検索する超並列計算機とからなり、超並列計算機の複数の異なるプロセッサ上に複数の異なる検索条件式を別個に設定し、記事データを全文検索して前記検索条件式に合致した結果をクライアント装置に送信するようにしたので、顧客によるカテゴリやキーワードの設定が不要で、顧客が意識することなく、顧客の興味を自動的に抽出し、即時に顧客の興味のある記事を提供することができるシステムが得られる。

【0071】

請求項 2 に記載の発明によれば、請求項 1 の効果に加えて、記事データは、サーバ装置とは別個の外部の記事生成コンピュータから受信したものであるもので、既存の記事配信機関が運営する記事配信システムの記事データの情報源として活用できるため、新たな記事データベースを構築する時間と費用を削減することができる。また、既存の記事配信機関がこの発明を採用した場合には、顧客に対して新たな付加価値を提供することができる。

【0072】

請求項 3 に記載の発明によれば、請求項 1 又は 2 の効果に加えて、顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として内容が類似する複数の記事を複数回に渡って顧客がアクセスした場合に、内容が類似する複数の記事に含まれる自然語を検索条件式の検索キーワードとして使用するようにしたので、顧客が検索条件を指定しなくとも興味ある記事のみが自動的に配信される。

【0073】

請求項 4 に記載の発明によれば、顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として内容の異なる複数の記事に顧客がアクセスした場合に、一定記

事数又は一定期間にアクセスした記事数を基準記事数と定め、該基準記事数における内容の異なる複数の記事に現れる同一の自然語を検索条件式の検索キーワードとして使用することとしたので、顧客が検索条件を指定しなくとも興味ある記事のみが自動的に配信される。

【 0 0 7 4 】

請求項 5 に記載の発明によれば、請求項 1 又は 2 の効果に加えて、顧客のアクセス履歴を顧客別に蓄積し、該アクセス履歴として顧客が特定の記事データを指定した場合に、該特定の記事データに含まれる自然語を検索条件式の検索キーワードとして使用するようにしたので、顧客が検索条件を指定しなくとも興味ある記事のみが自動的に配信される。

【 0 0 7 5 】

請求項 6 に記載の発明によれば、請求項 4 の効果に加えて、検索条件式は、顧客が指定した特定の記事データにおける自然語の出現回数と顧客のアクセス履歴中に含まれる複数の記事中に自然語が出現する割合とによる重み付けをした特徴的な自然語を検索キーワードに採用したので、顧客は極めてノイズの少ない記事を入手することができる。

【 0 0 7 6 】

請求項 7 に記載の発明によれば、請求項 1 乃至 6 のいずれか 1 つの効果に加えて、クライアント装置が携帯端末装置であるから、記事配信を受ける時と場所を選ないため、顧客は必要な時に必要な場所で最新の興味記事を入手することができる。

【 0 0 7 7 】

請求項 8 に記載の発明によれば、クライアント装置と該クライアント装置からの要求に応じて情報を提供できるサーバ装置とがネットワーク接続されている場合の記事配信方法において、クライアント装置を使用する顧客のアクセス履歴を分析して複数の異なる検索条件式を生成し、該複数の異なる検索条件式を複数の異なるプロセッサを有する超並列計算機の複数の異なるプロセッサ上に別個に設定し、外部の記事生成コンピュータから逐次送出される記事データを複数の異なる検索条件式で同時並行的に全文検索し、検索条件式に合致した結果をクライア

ント装置に提供するようにしたので、顧客によるカテゴリやキーワードの設定が不要で、顧客が意識することなく、顧客の興味を自動的に抽出し、即時に顧客の興味のある記事を提供することができる。

【図面の簡単な説明】

【図 1】 この発明の実施の形態に係る興味記事配信システムの構成を示すブロック図である。

【図 2】 同実施の形態に係る興味記事配信システムのサーバ装置に含まれる超並列計算機の内部処理の仕組みを示した模式図である。

【図 3】 同実施の形態に係る興味記事配信システムを使用する場合の手順を示すフローチャートである。

【図 4】 同実施の形態に係る興味記事配信システムのクライアント装置の画面イメージ図である。

【図 5】 同実施の形態に係る興味記事配信システムにおける、指定された記事内容から検索条件式を自動生成する手順を示すフローチャートである。

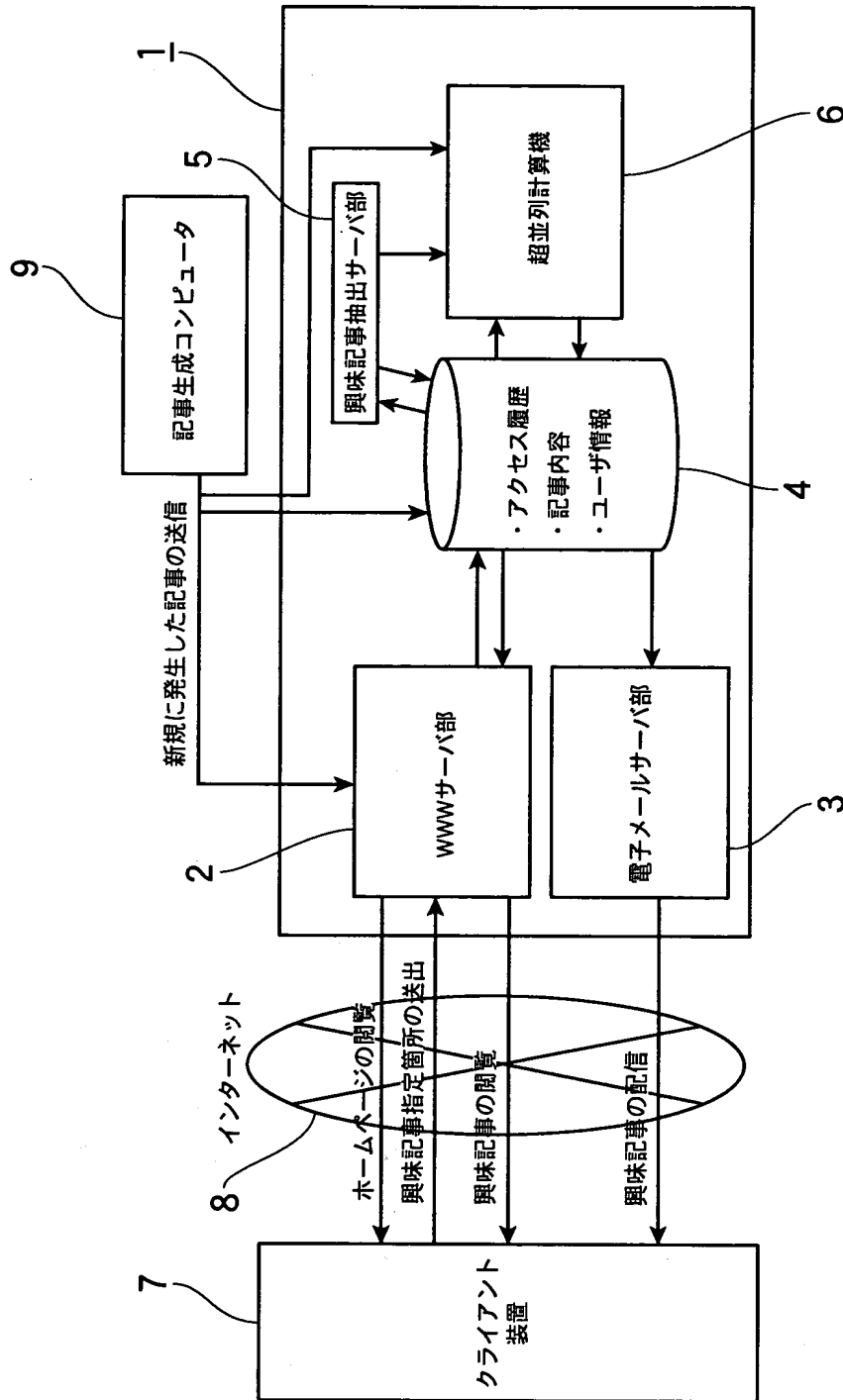
【図 6】 同実施の形態に係る興味記事配信システムにおける、アクセス履歴の統計処理及びこれに基づく検索条件式の自動生成の手順を示すフローチャートである。

【符号の説明】

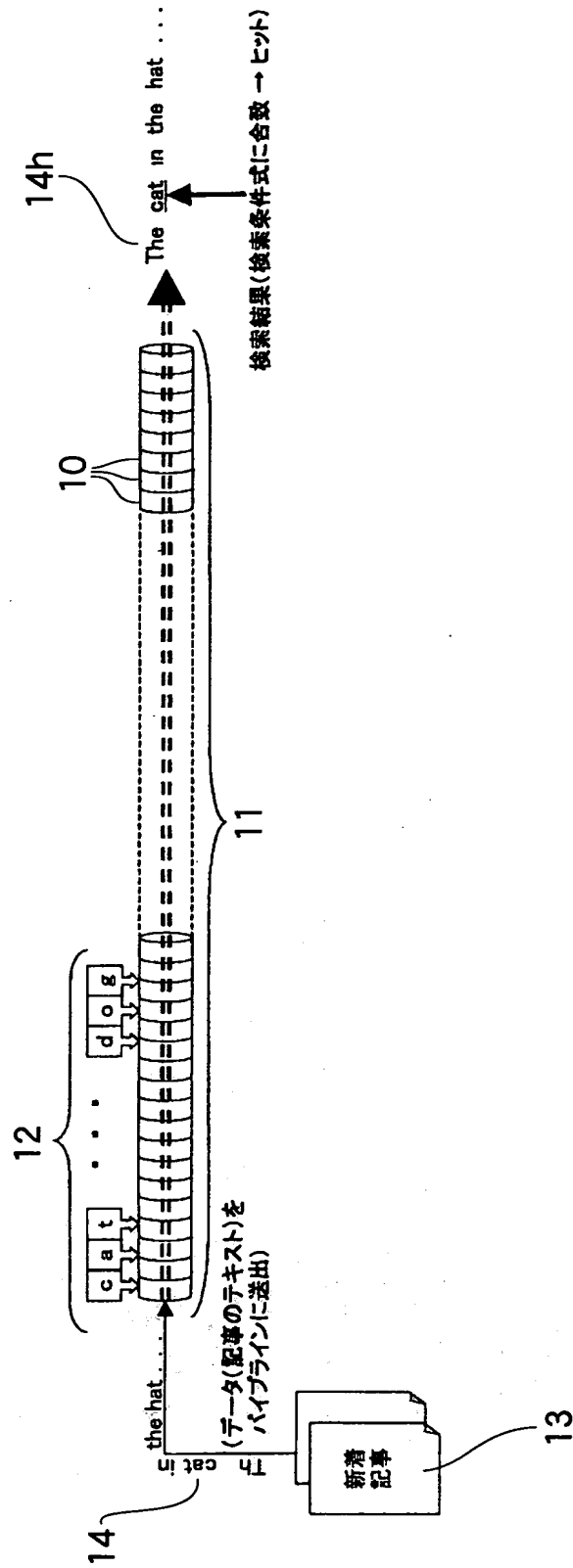
- 1 サーバ装置
- 2 WWWサーバ部
- 3 電子メールサーバ部
- 4 データベースサーバ部
- 5 興味記事抽出サーバ部
- 6 超並列計算機
- 7 クライアント装置
- 8 インターネット
- 9 記事生成コンピュータ

【書類名】 図面

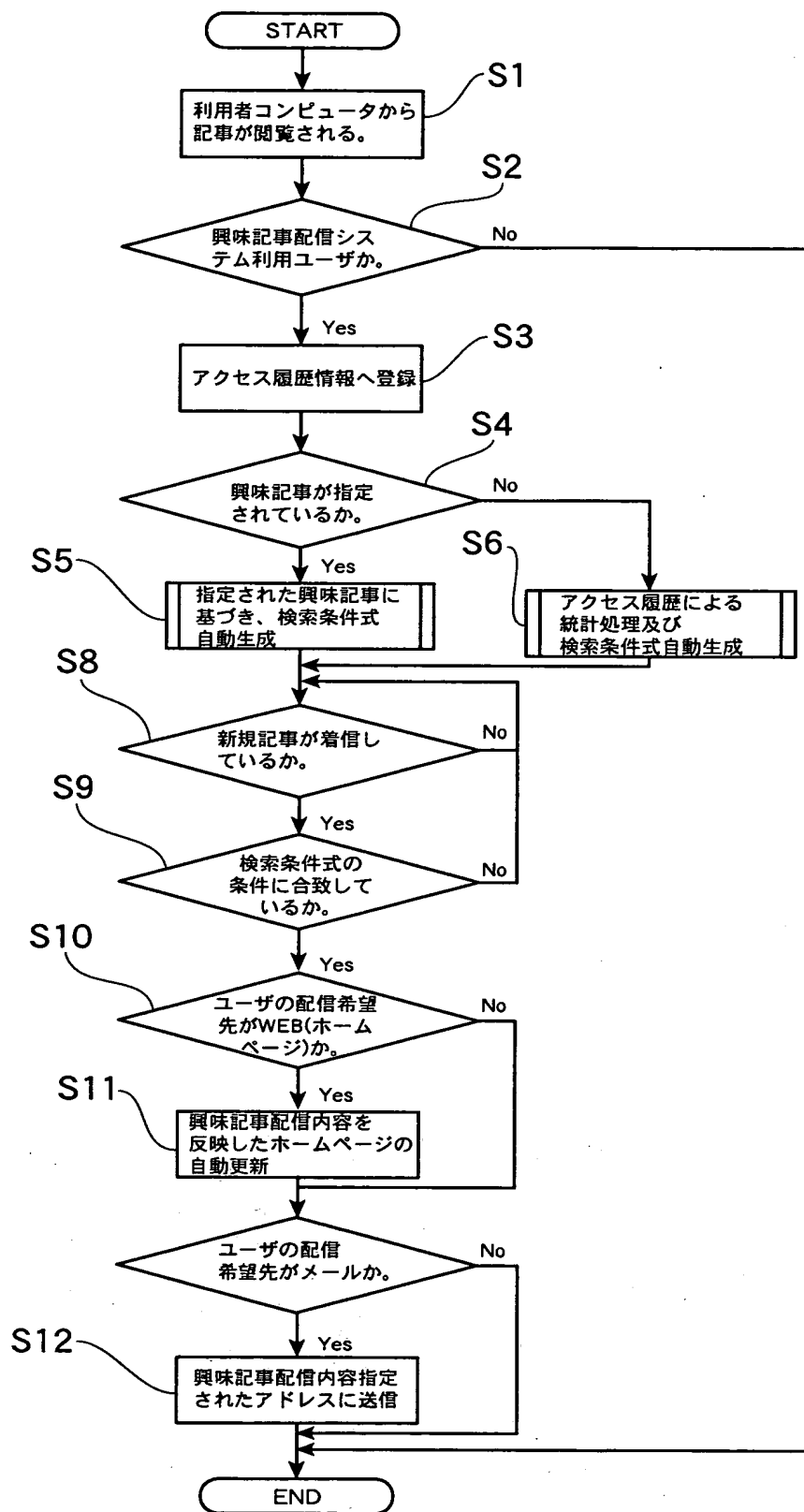
【図 1】



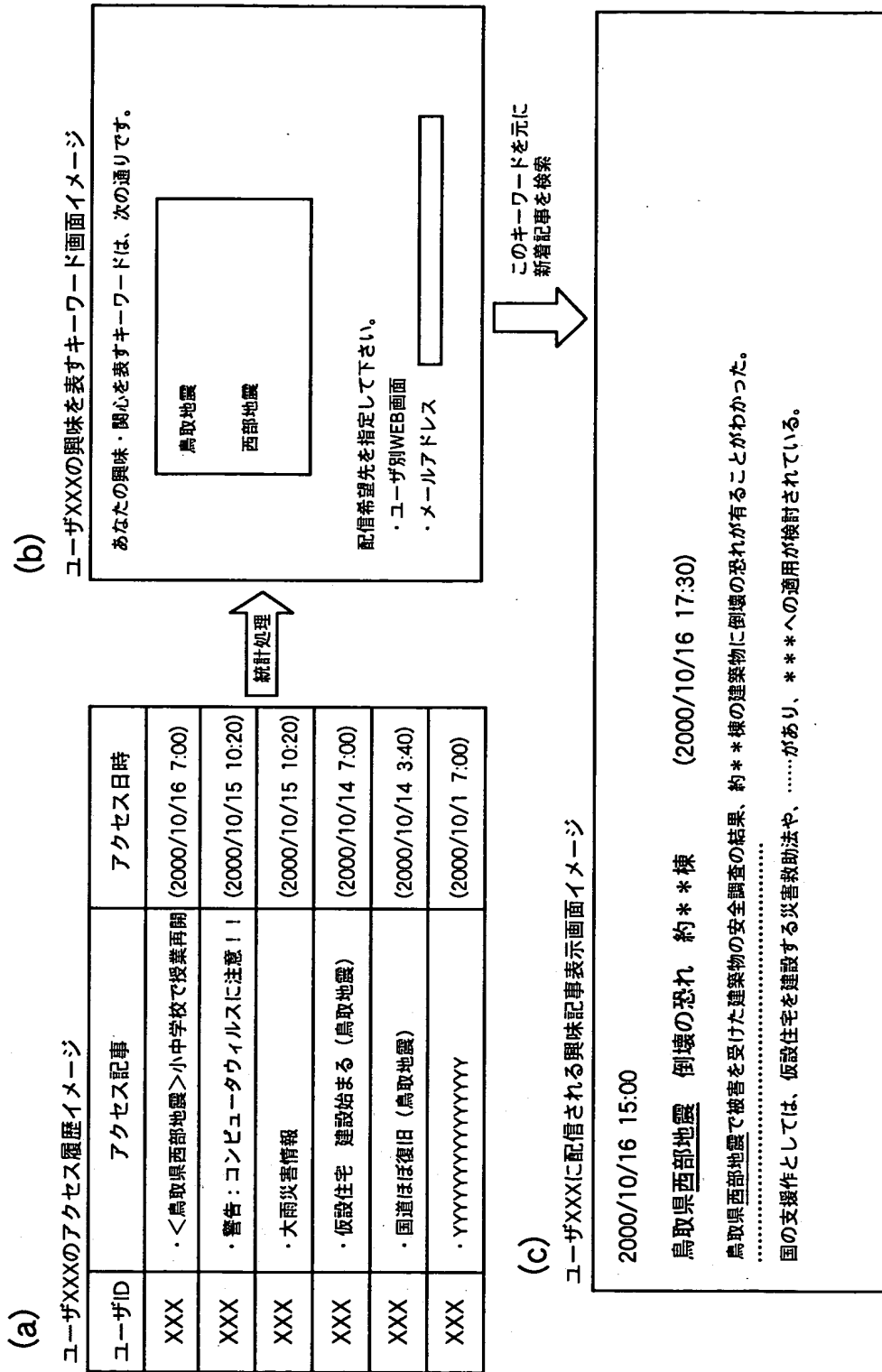
【図 2】



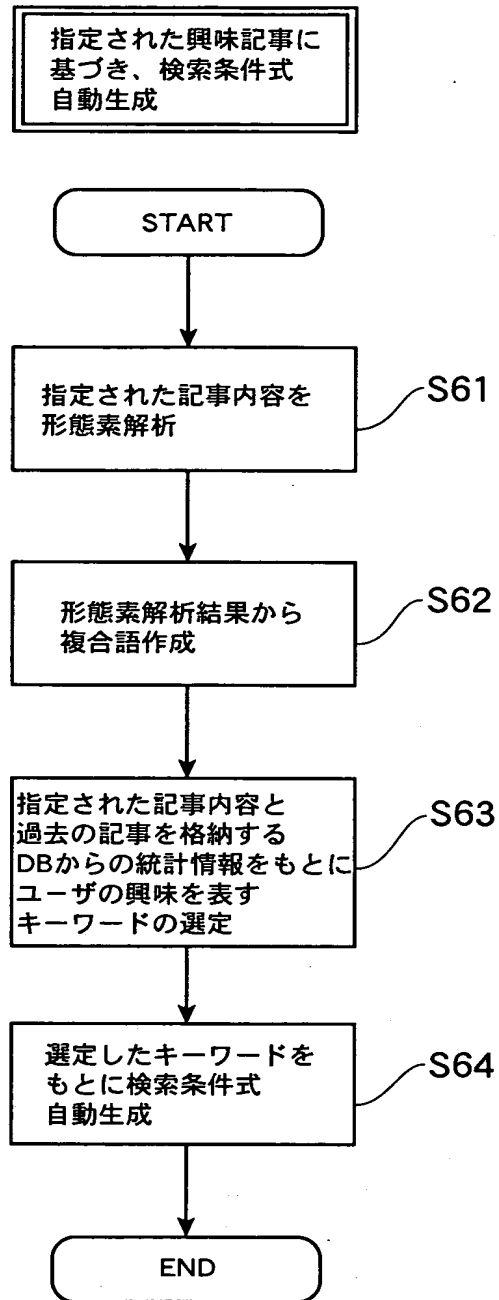
【図 3】



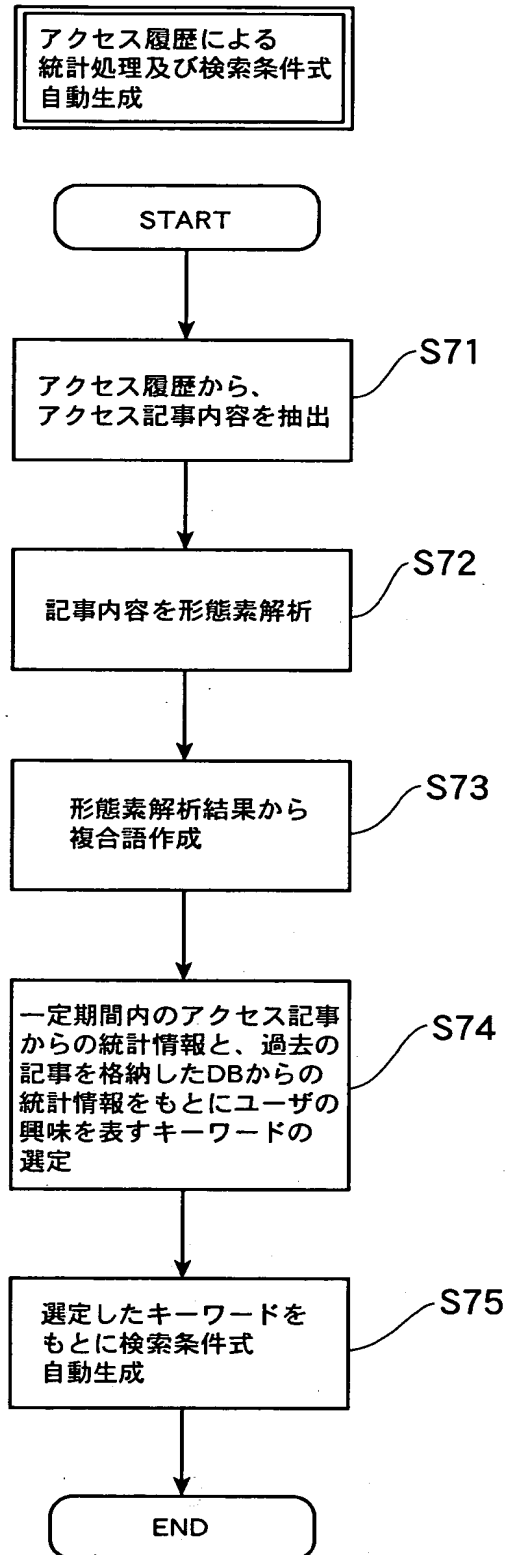
【図 4】



【図 5】



【図 6】



【書類名】 要約書

【要約】

【課題】 顧客によるカテゴリやキーワードの設定が不要で、顧客が意識することなく、顧客の興味を自動的に抽出し、即時に顧客の興味のある記事を配信するシステムとその方法を提供する。

【解決手段】 サーバ装置 1 は、クライアント装置 7 からの要求を受けこれに応答する WWW サーバ部 2 と、クライアント装置 7 を使用する顧客の個人情報及びアクセス履歴を保存するデータベースサーバ部 4 と、該データベースサーバ部 4 に保存された顧客のアクセス履歴を分析して検索条件式を生成する興味記事抽出サーバ部 5 と、該興味記事抽出サーバ部 5 で生成された検索条件式に基づいて、外部の記事生成コンピュータから逐次送出される記事データを検索する超並列計算機 6 とからなり、超並列計算機 6 の複数の異なるプロセッサ上に複数の異なる検索条件式を別個に設定し、サーバ装置 1 が保存している記事データを複数の異なる検索条件式で同時並行的に全文検索し、検索条件式に合致した結果をクライアント装置 7 に提供する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願2000-351144
受付番号	50001486813
書類名	特許願
担当官	第七担当上席 0096
作成日	平成12年11月20日

<認定情報・付加情報>

【提出日】	平成12年11月17日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [591102095]

1. 変更年月日 1991年 5月16日
[変更理由] 新規登録
住 所 東京都港区浜松町2丁目4番1号
氏 名 三菱スペース・ソフトウェア株式会社